

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
2 August 2001 (02.08.2001)

PCT

(10) International Publication Number  
**WO 01/54480 A2**

(51) International Patent Classification: Not classified

(21) International Application Number: PCT/US01/03214

(22) International Filing Date: 31 January 2001 (31.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/494,412 31 January 2000 (31.01.2000) US

(71) Applicant: ENGAGE TECHNOLOGIES [US/US]; 100  
Brickstone Square, Andover, MA 01810 (US).

(72) Inventor: JAYE, Daniel, B.; 650D Brookside Drive, An-  
dover, MA 01810 (US).

(74) Agents: KELLY, Edward, J. et al.; Foley, Hoag & Eliot  
LLP, One Post Office Square, Boston, MA 02109 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— with declaration under Article 17(2)(a); without classifica-  
tion and without abstract; title not checked by the Interna-  
tional Searching Authority

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.



**WO 01/54480 A2**

(54) Title: SYSTEM AND METHOD FOR INFERRING DEMOGRAPHIC PROFILES

(57) Abstract:

BEST AVAILABLE COPY

- 1 -

## SYSTEM AND METHOD FOR INFERRING DEMOGRAPHIC PROFILES

Field of the Invention

5 This application relates to the field of user behavioral data and more particularly to the field of Internet browsing habits and the purchasing and demographic user characteristics.

Background of The Invention

10 Users connected to the Internet can access an ever increasing number of Web sites to obtain information or to conduct business. Each Web site has associated therewith a unique identifier, that can be represented as a URL (Uniform Resource Locator). The user can connect to the site by, for example, typing the URL (such as "www.yahoo.com") or selecting a site from a predefined menu. Once at the site, the user can review the content presented on the site, and can provide information and instructions for directing the site to provide services and goods.

15 The providers of such commercial Web sites are generally interested in the demographic profile or attributes of a user in order to be better able to target advertising, products and/or services to specific users or groups of users. In many cases, the user may provide his or her demographic profile by responding to prompts or by filling out a registration form. The user demographic profile may include gender, age, marital status, education, profession, income and/or the geographic region as indicated, for example, by the user's ZIP code. A number of Web sites providing information, goods and services may attract a cross-section of users sharing at least demographic characteristics. For example, the Web site "www.bettycrocker.com" may be preferred by women between the ages of 30 and 55 and having an interest in the culinary arts. Conversely, the Web site "www.harley-davidson.com" may preferably be visited by men between the age of 35-44 interested in the outdoors. In other words, Web sites may be frequented by many different users, many of which share certain demographic characteristics, leading to a distribution of user profiles.

20  
25  
30 Frequently, however, a user may be reluctant to provide detailed personal information, in which case the Web site providers have employed other means to generate

- 2 -

an approximate profile of the user, for example, based on the user's click stream.

Optionally, these techniques may be employed in a way that refrains from associating user identity information, such as name and address information, with information descriptive of the user's profile. Although these profiling techniques may work well, the problem with this approach is that the provider basically has to start with a "blank" sheet, i.e., zero data, for each new user and build the profile from sometimes random click stream information. This process may be slower than desired. Thus, a web site that experiences sporadic but intense user contact, such as a web site that sells toys during the Holiday season, may need profile information more quickly than presently available.

It is therefore desired to have that methods and systems for providing a demographic user profile more rapidly, in particular for new users who have not visited a Web site before. It is further desirable to provide a measure of the demographic user profile based on the Web site visited by the user and the user's click stream.

## Summary of the Invention

The systems and methods described herein include systems and methods for inferring the demographic properties of users visiting a web site. In one aspect, systems are provided wherein a set of discriminating web sites are identified. These discriminating web sites are identified by examining the browsing activity of certain known users that have associated with them profile information that is representative of demographic information for these users. For these known users a set of discriminating web sites can be identified, wherein a discriminating web site is understood as being prototypical of the characteristics of these known users. The demographic information of these discriminating web sites may be employed to infer the demographic properties of other users visiting these web sites.

More specifically, the invention provides methods for generating a demographic profile for an unknown user, comprising recording computer activity of the unknown user in response to the information provided to the user by at least one of the digital processors, and combining the recorded computer activity of the unknown user with a computer demographic score of the at least one of the digital processors, the computer demographic score being based on demographic information obtained from known users, to generate the demographic profile of the unknown user.

- 3 -

Further, the invention may provide a process for generating a demographic profile of an unknown user accessing a server having a server profile. The process may include recording computer activity by the unknown user in response to information provided by the server; determining whether the recorded computer activity by the unknown user is greater than a predetermined activity value, and combining the recorded computer activity by the unknown user with the server profile to form the unknown user demographic profile. The recorded computer activity by the unknown user can be checked to determine if it is less than a predetermined activity value, and if so, the process may set the unknown user demographic profile equal to the server profile.

10 In an optional practice a weighting function may be applied to the recorded computer activity by the unknown user based on a duration of the computer activity. The applied weighting function may be selected to reduce the significance of a computer activity having a long duration.

The systems and methods described herein thus provide a process for generating a user demographic profile of an unknown user accessing at least one server. These processes may comprise identifying a user accessing the at least one server and recording user activity on the server, and determining, as a function of the user identification, whether the user is an unknown user or a known user. For an unknown user, the process may monitor at least a duration of the user activity and assign a demographic score to the unknown user based on the monitored user activity and a server profile of the at least one server accessed by the unknown user. The demographic score may be combined with an existing demographic score of the unknown user, and the demographic profile of the unknown user may be set equal to the combined demographic score.

25 In a further aspect, the invention may be realized as computer programs having instructions for causing a computer to record computer activity of a user responding to information provided by at least one of the computers. The program may identify the user as one of a known or an unknown user. The program may compare the computer activity, for an unknown user, with a predetermined activity and assign to the unknown user a demographic score which is based on the computer activity and a computer demographic profile characteristic of the computer if the computer activity exceeds the predetermined activity, and on the computer demographic profile alone, if the computer activity is less

than the predetermined activity. The program may then combine the demographic score with another existing demographic score for the same unknown user generated during a previous session by the unknown user with the same computer or with another computer; and provide from the combination of the demographic scores a user demographic profile of the unknown user.

Further features and advantages of the present invention will be apparent from the following description of certain illustrated embodiments and from the claims.

#### Brief Description of the Drawings

- FIG. 1 is a functional block diagram a computer network;
- FIG. 2A is a flow diagram of a process for inferring an unknown user profile according to the invention;
- FIG. 2B is a flow diagram of a process for updating a URL profile according to the invention;
- FIG. 3 is a data flow diagram of the process of FIGS. 2A and 2B;
- FIG. 4 shows the organization of a hash table for known users;
- FIG. 5 is a flow chart for computing URL profiles of known users;
- FIGS. 6A-6C and 7 are detailed flow charts of the process of FIG. 5;
- FIG. 8A and 8B show the organization of a hash table for unknown users; and
- FIG. 9 is a flow chart for computing URL profiles of unknown users.

#### Detailed Description of Certain Illustrated Embodiments

To provide an overall understanding of the invention, certain illustrative embodiments will now be described. However, it will be understood by one of ordinary skill in the art that the systems described herein can be adapted and modified to provide systems for other suitable applications and that other additions and modifications can be made to the invention without departing from the scope hereof.

A demographic profile of an unknown user (hereinafter referred to as "unknown user profile") interacting with one or more computers, such as Web servers, may be compiled based on the computer activity of the unknown user and a computer demographic profile (hereinafter referred to as "URL profile") associated with the computer or Web

server accessed. The URL profile may be derived from and updated in response to the browsing activity of known users and/or from demographic information provided by the known users. Known users can be distinguished from unknown users by the cookies exchanged between the user's PC and the Web server. The demographic profile of the  
5 known users can be based on their activity on several Web servers.

To provide a better understanding of the invention, certain terms should first be defined. A complete set of demographic information may be available for some users. These users will hereinafter be referred to as "known" users. The browsing activity of the known users is obtained from the Web sites (hereinafter referred to as "URL") they have  
10 visited over a period of time along with the number of visits (hits) and the browsing duration for each URL. This information is used to identify certain URLs, which are prototypical of the characteristics of these users. Such URLs are referred to as "discriminating" URLs. In contrast, users whose demographic characteristics are not known are referred to as "unknown" users. The process and system of the invention  
15 attempts to establish a reliable demographic profile of an unknown user based the browsing activity of the unknown user and the demographic properties of the discriminating URLs visited.

A measure in the form of a demographic score vector (dscore) is assigned to each URL. Each element of this vector represents an attribute value of the demographic  
20 information. This dscore element is a representation of the probability that a user visiting this URL can be associated with that particular attribute value. The demographic profile of a user can also be expressed as a dscore vector.

Typically, the dscore vector for either the user or the URL may include the following data:

25 Gender (male/female)  
Age (2-11, 12-17, 18-24, 25-34, 35-44, 45-54, 55+)  
Marital Status (Single, married, divorced)  
Children (by age: i.e., 1+ children 0-5, 1+ 6-11, 1+ 12+)  
Education (some high school, high school degree, some college, college degree,  
30 advanced degree, professional degree)

- 6 -

Profession (broad categories, e.g.: Technical, Professional, Homemaker, student, Trade, Management, Sales and Marketing, Service).

Income (US\$ < \$25k, \$25k-\$35k, \$35k-\$45k, \$45k-\$75k, \$75k-\$100k, >\$100k)

Geography (Country, State, Town, ZIP code or Area code).

5 The broad categories are the attributes, and the sub-categories enclosed in parentheses are the attribute values.

An exemplary URL dscore may have the form shown in Table 1:

User	Age Group			Gender		Marital Status			Others
	0-20	20-40	>40	M	F	S	M	D	
U1	0.2	0.7	0.1	0.4	0.6	0.3	0.5	0.2	

As seen in Table 1, a typical user logging on to the exemplary Web site is expected  
 10 to be a married women between the age of 20 and 40 years. It is therefore not unreasonable to expect that an unknown user logging on to the same Web site will have a statistically significant probability to fit this profile.

Referring now to FIG. 1, a part 10 of a computer network includes a profile server 12, personal computers (PC) 14, 15, and Web servers 16 and 17. The PCs 14 and 15 may  
 15 be any one of a variety of conventional, commercially available, hardware and software combinations configured to access Internet servers by any one of a variety of suitable means. Similarly, the Profile Server 12 and the Web servers 16 and 17 may also be any one of a variety of conventional, commercially available, hardware and software combinations configured to provide conventional Internet services to users. In some instances, such as  
 20 those described below, the conventional server software is supplemented to provide the functionality discussed herein. The PCs 14, 15 and the servers 12, 16 and 17 communicate with each other via communication links 22, 24-27 which are all connected to a communication channel 21, such as the Internet.

For the system described herein, each of the servers 16, 17 has a unique  
 25 identification, generally referred to as a URL (Uniform Resource Locator) serving as the server's network address. The Web server 16 may, for example, provide advertisements to PC 14 or 15 accessing the server 16. Alternatively, the Web server 16 may provide the advertisements, for example, in banner form to another server 17 with which the PC 14 is

communicating. The Web servers 16, 17 may also be used in electronic commerce applications offering goods and services to the users 14, 15, as is known in the art. One of the servers 12 is designated as a "Profile Server" 12 capable of monitoring user interactions with the Web servers 16, 17. The Profile Server 12 can communicate with one or more  
5 Web servers 16, 17 which run profiling software that allows the Web servers 16, 17 to monitor and record user input in the PCs 14, 15, such as the users' click stream. The Profile Server 12 may monitor the user input in real time or download user input data offline. The Web servers 16, 17 can also query the Profile Server 12 to obtain information about user interaction with their own Web server 16 and 17, respectively, or with another  
10 Web server or other Web servers. The Profile Server 12 can control and limit the user information made available to the Web server 16 or 17 from another Web server.

Referring now to FIG. 2A, in process 30 an unknown user logs on to a Web site, step 32. The unknown user is most likely drawn to this Web site because of its displayed or assumed content. The Web server of the Web site returns to the user's PC a cookie ID  
15 containing at least the URL of the Web server and a time stamp recording the time and duration of the user's click on information provided by the Web server, step 34. A "cookie" is generally referred to as a packet of information sent by a server to a browser and then returned by the browser to the server each time the browser accesses the server. Cookies may contain any information the server chooses and are used to maintain state  
20 between otherwise stateless transactions, such as HTTP transactions. Typically this is used to authenticate or identify a registered user of a web site without requiring the user to sign in again every time he/she accesses that site. Other uses are, e.g. maintaining a "shopping basket" of goods selected for purchase during a session at a site, site personalization (presenting different pages to different users), and/or tracking a particular user's access to a  
25 site.

The software running on the visited Web server records the unknown user's browsing activity on that Web site, step 36. The browsing activity is likely to reflect the user's interests and can be used to compile a user's interest score (iscore) which may be retained in a database of the Profile Server 12 or the visited Web server, where the iscore  
30 can be updated when the user logs on to the Web site again. The compilation of interest



scores is described, for example, in the above mentioned and commonly assigned US patent application entitled "SYSTEM AND METHOD FOR BUILDING USER PROFILES".

The software computes the demographic profile of the unknown user (unknown\_user\_dscore) based on the user's browsing activity monitored in step 36 and the dscore of the visited Web site (URL\_dscore), step 38. The URL\_dscore is obtained based on information of known users having an established user profile for the visited Web site using process 50, which will be described below with reference to FIG. 2B. The unknown\_user\_dscore is then stored in a history table to be used for a subsequent login of the same unknown user, step 42.

Referring now to FIG. 2B, process 50 computes and updates the URL\_dscores of discriminating Web sites (URLs) based in the browsing activity of known users. A known user logs on to a discriminating Web site, step 52, and the user PC sends a cookie with user information from prior interactions to the Web server, step 54. The software running on the visited Web server records the known user's browsing activity on that Web site and updates a browsing log of the known user, step 56. The updated browsing log is then compared with existing records of known users. If the updated browsing log is statistically identical to the existing records, then the URL\_dscore of the Web site remains unchanged. If, on the other hand, the updated browsing log is statistically different from the existing records, then the URL\_dscore of the Web site is updated to reflect the altered browsing patterns of known users, step 58. The updated URL\_dscore can then be stored in a file or database, step 60, and is used to compute the unknown\_user\_dscore in step 38 of FIG. 2A, as described above.

Referring now to FIG. 3, the processes 30 and 50 of FIGS. 2A and 2B will now be described with reference to a data flow diagram 70. The software running on, for example, Web server 16 which may be monitored by the Profile Server 12, receives a user's cookie, the URL visited and a time stamp indicating, for example, the time of the visit and the duration of the browsing activity, step 62, creating a log record. The log record is processed by log record processing means, step 64, wherein the users are segregated according to the received cookies into known users for which a demographic profile has been established, and unknown users for which a dscore is to be computed using the system

and method of the invention. Interest scores (iscores) based on the user browsing activity may also be compiled in step 64, as mentioned above.

The records 78 of known users will be used for the URL\_dscore computation 80 to update a URL file of known users 82 based on the new browsing activity of the known users. The process 70 also maintains a database 84 with the dscores of other Web sites (URLs) profiled by the Profile Server 12. The database 84 is also updated as needed or on a periodic basis. The updated URL dscores are then supplied to the log record processing means 64 which computes a session dscore of the unknown user for the current session on the visited Web site (URL) based on the user's browsing activity, in particular the browsing duration, and the URL\_dscore of the visited Web site, step 66. The session dscore is then merged with dscores compiled for the unknown user during previous sessions and stored in a history file 74, providing a new dscore of the unknown user, step 72. The dscore and other browsing attributes, such as the URLs of the visited Web sites and browsing duration, in the history file may also be updated. The history file of an unknown user may have the following form:

Cookie	Iscore list	Dscore list, Duration count	Last-calc-date
--------	-------------	-----------------------------	----------------

Referring now to FIG. 9, a flow chart 200 describes the browsing activity of an unknown user. First, the time duration for which the unknown user browses a Web site (URL) is recorded, step 202. Duration counts have proven to represent a better measure for calculating an unknown user's dscore than merely the number of visits (hits). Storing all browsing activity, including all hits, for all visits of unknown users to all URLs is computationally extensive, since both the number of unknown users and discriminating URLs can be very large.

Moreover, only those duration values which are significant enough to indicate the interest of the user in that URL should be taken into consideration. However, undue importance should not be given to a single visit having a large duration. For this reason, a non-linear function is applied to the duration values which caps large duration values, step 204 (see Eq. 12 in the Appendix). The value of this non-linear function is selected to increase rapidly for small but significant values of duration, but remains constant for large duration values. The non-linear function also weights the discriminating ability of the URL

browsed so that the dscores of the URLs that are more discriminating, are scaled by a larger amount.

If it is decided in step 206 that the browsing activity of the user is "significant", i.e. more extensive than a predetermined threshold value, then the duration count (DC) is computed and updated, step 210, using Eq. 2 of the Appendix, and the dscore of the unknown user for the URL browsed is computed for each attribute value from the weighted and aged duration count and the URL dscore of that URL, step 212. The duration counts are "aged" in a manner known in the art to reduce the effect of old duration counts on the new dscore values.

At the end of a session, the dscore values of all URLs browsed by the unknown user are computed, step 214, using Eqs. 12 and 13 given in the Appendix. The dscore value of the unknown user will be more heavily weighted towards the more discriminating URL (see Table A5 in the Appendix). The new dscore values computed in step 214 are then merged with the old dscore values of the unknown user stored in the history table 74, step 216.

Before merging, the old duration counts and the old dscore values are first aged, using Eqs. 11-18 listed in the Appendix. The merged and aged dscore values represent the current dscore values of the unknown user, step 218.

Alternatively, if it is determined in step 206 that the unknown user's browsing activity is not "significant" by being smaller than the predetermined value, then the predicted dscore for the unknown user is set equal to a typical user profile for the URL browsed, step 208. In other words, the demographic profile of the user showing insignificant browsing activity will be set to the profile shown, for example, in Table 1 if the user accesses this exemplary Web site.

As mentioned above, known users may be distinguished from unknown users based on the cookie attributes. Log processing process 64 has a list of cookies of the known users. If the cookie exists in the list, all records in the session will be identified as those of known users. This comparison could be made more efficient by hashing the cookies as illustrated in FIG. 4.

Referring now to FIG. 4, a hash table 90 may be implemented in the form of an array of pointers 62. All pointers in the hash table initially have NULL values. When a known user cookie ID 94, 96, 98 hashes to a hash bucket 96', 98', the array is indexed by

bucket number and assigned a corresponding pointer, thereby linking the cookie to the pointer. To resolve conflicts between two pointers 94, 96 pointing to the same hash bucket 96', the cookies 94, 96 will be chained. The number of hash buckets 96', 98' may be set equal to the number of known users.

5           The hash table 90 of known users may, in some embodiments, always be present in a memory. An incoming hashed cookie is checked whether it points to an existing hash bucket. With a suitable hash function and a suitable number of buckets, known and unknown users may be identified after a very few comparisons.

          For computing the URL\_dscores in step 80, all incoming known user records are  
10       first sorted according to the URLs visited. If the log records of the known user are stored in more than one file, each file will be sorted individually with respect to the URLs. The sorted files will then be merged and organized for each URL in, for example, ascending order. These output files are advantageously read in the same order in which they were created.

15           Referring now to FIG. 5, a flow chart 100 shows the process step 80 of FIG. 3 for computing the URL\_dscores for known users. First, a duration count (DC) is computed for each user and each visited Web site (URL), step 102. The duration count represents the time during which a user interacts with the URL. From the duration count, an activity count is computed which is then "aged" to reflect the elapsed time since the user's last  
20       access to the URL, step 104. The new dscore of the URL for known users is then computed from the aged activity count, step 106. The URL\_dscores are then updated and the temporary files created during the process 100 are deleted, step 180.

          FIGS. 6A - 6C illustrate the process 100 of Fig. 5 in more detail. Referring first to FIG. 6A, the read pointer of the first record in the known user log 78 and of the first file in  
25       the URL\_known user file 82 are initialized, step 112. As mentioned above, the URL\_known user file 82 is already sorted according to the URLs, for example in ascending order. The first record in the known user log 78 is read, step 114, and that URL is compared with the URL in the current record of the URL\_known user file 82, step 116. Depending upon whether the URL from the user browsing log record is less than, equal to, or greater  
30       than the URL from the record of the URL\_known user file 82, the one of the following three process steps is executed:

- 12 -

If it is determined in step 116 that the URL in the user browsing log record is less than the URL in the URL\_known user matrix file 82, a buffer for storing duration counts of all users is allocated and initialized to zero, step 118. The records for that URL are read one by one from the user browsing log record, step 120, and the duration is normalized (see  
5 Eq. 1 in the Appendix) to give duration counts, step 122. The duration counts are then aged up to the current date-time (Eq. 5), step 124, and added to the duration counts of that user in the buffer (Eq. 2), step 126. After all log records for that URL are processed, the duration counts of all known users are normalized to produce activity counts (Eq. 3), step 128. The dscores for the respective URL are computed using the known user dscores and  
10 the accumulated activity counts (Eq. 7), step 130. If the dscores so computed show that the URL is discriminating, then the dscores of the URL are appended to and inserted in a temporary file 'insert\_file', and the activity counts for all users for that URL are also appended to a new file of the URL\_known user file 82 in form of a 'newUrlAcFile' along with a bit denoting whether that URL is discriminating or not, step 132. As mentioned  
15 above, URLs that track the browsing activity of known users over a period of time along with the number of sites and duration of browsing for the URL are referred to as discriminating URLs.

If it is determined in step 116 that the URL in the user browsing log record is equal to the URL in the URL\_known user file record, then the activity counts for all known users  
20 for that URL are read from the URL\_known user file 82 into a memory buffer, step 140. The activity counts are denormalized to get duration counts, step 142. The duration counts are then aged up to the current date-time, step 144, and the aged duration counts are added to the duration counts of that user in the memory buffer, step 146. If it is determined in step 148 that the URL was previously discriminating and is now no longer discriminating  
25 (as is denoted by the discriminating status bit in the URL\_known user file record), then the duration counts are appended to a temporary delete file 'del\_File', step 150. Otherwise, the process 100 goes to step 152. If it is determined in step 152 that the URL was previously discriminating and still remains discriminating, then the URL dscore is computed as in steps 128 and 130 and appended to a temporary update file 'update\_file', step 154. If, on  
30 the other hand, it is determined in step 152 that the URL was not discriminating before and has now become discriminating, then the duration counts are appended to a temporary file

- 13 -

'insert\_file', and the activity counts for all users for that URL along with the updated discriminating status bit are also appended to the 'newUrlAcFile', step 132.

If it is determined in step 116 that the URL in the user browsing log record is greater than the URL in the URL\_known user matrix record, then the activity counts for all known users for that URL are read from the URL\_known user matrix record into a memory buffer, step 160. The activity counts are denormalized to get duration counts, step 162. The duration counts are then aged up to the current date-time, step 164, and new activity counts and URL dscores are computed, step 166. If it is determined in step 168 that the URL was previously discriminating and is now no longer discriminating (as is denoted by the discriminating status bit in the URL\_known user matrix record), then the activity counts are appended to a temporary file 'del\_file', step 170. The URL dscore and the associated new activity counts along with the current discriminating status bit are appended to the 'newUrlAcFile', step 172. If, on the other hand, it is determined in step 168 that the URL was and still is discriminating, then the URL dscore is appended to a temporary 'update\_file', step 154.

Referring now to FIG. 7, process 180 next updates the URL dscore database 84 and the URL\_known user file 82. After both the input files, i.e., the known user browsing log 78 and the URL\_known user file 82, have been processed in the manner described above, the updated URL\_known user file 82 now resides in the 'newUrlAcFile'. The other files produced by the process 100 are the temporary del\_file, update\_file and insert\_file. The del\_file is processed first, and all non-discriminating URLs are deleted from the URL demographics file, step 182. The update\_file is processed next, step 184, wherein the 'newUrlAcFile' updates the URL\_known user file. The insert\_file is processed last, step 186, where all new discriminating URLs are inserted. The three temporary delete, update and insert URL dscore files are then deleted, step 188. Finally, the old URL\_known user file is deleted, step 190, and the 'newUrlAcFile' is renamed as the new URL\_known user matrix file, step 192. The processes 100 and 180 maintain the sorted order of the URL\_known user file.

Referring now to FIGS. 8A and 8B, the computation speed of dscores and duration counts of unknown users can be increased by providing a memory cache. The main components of the memory cache are a hash table 220 and an aging queue. The hash table

220 is similar to the hash table 92 for known users. The URL, such as URL 1, is hashed to provide the bucket number 222' of the hash bucket. The hash bucket 222' contains a pointer to the chain of URLs 222, 224 that hash to that hash bucket 222'. For example, the URL 1 and 2 and the associated dcsore hash to the same bucket 222'. The URLs can be distributed uniformly across all hash buckets by selecting a suitable hash function.

The aging queue is provided to identify URLs that can be replaced, because they have not been accessed for some time. The memory cache may, for example, be already full when a URL is fetched from the database to be stored it in the memory cache. A queue of all URLs is then formed in the memory cache. Each URL has an aging queue pointer 230 pointing to the previous URL in the chain and an aging queue pointer 232 pointing to the next URL in the chain. A newly fetched URL is added to the tail of the queue. In the example illustrated in FIG. 8A, the previous pointer of URL 1 (222) points to URL 2 (224), whereas the next pointer of URL 1 points to URL n. Likewise, the previous pointer of URL 2 (224) points to URL n (226), whereas the next pointer of URL 2 points to URL 1. The aging queue order is therefore URL 1, URL 2, URL n. When a URL, for example URL n, from the cache is accessed, the URL is moved from its current position and placed at the tail of the queue, as indicated by the arrows in Fig. 8B. Thus the URLs at the tail of the aging queue are the most recently accessed URLs, whereas those at the head of the aging queue have been accessed least recently. The URL at the head of the queue is replaced, if necessary.

In the foregoing, it is assumed that the known users provide all necessary attribute values of all attributes. However, some known users may choose not to supply values for some demographic attributes, such as their age or household income. Such values will be referred to as missing values. These missing values, however, should be incorporated to reduce the error in the URL\_dscores and thereby minimize the error in the demographic profile predicted for unknown users. In such cases, one of the following strategies may be adopted: The missing attribute value field may be replaced by NULL values which are ignored in the dscore computation. Alternatively, the missing demographic information for known users may be predicted either statically by creating an auxiliary function called a "bridging agent" in ML terminology that can predict the value of the missing attribute. Conventional data mining algorithms, such as the association rule finding algorithm; can be

applied to the existing known user demographics file and used to predict missing demographic information based on the demographic attributes from other known users. This may need to be done only once, with confidence values for the missing demographic attributes permanently stored in the known user demographics file.

5           Alternatively, the missing demographic information for a known users may be predicted dynamically by treating the known user as an unknown user and finding the dscore in a manner similar to the dscore computation for unknown users discussed above. This method would take into account the browsing patterns of the known user along with the demographic attributes which other known users have provided.

10           Each of the aforescribed methods for predicting the missing demographic information for known users has disadvantages. For example, if the unknown attributes values of the known users are discarded, then vital relationships between different attributes may be lost. Also, extra information will have to be recorded about the total activity count of known users who have not declared a particular attribute. This extra information will be  
15           required because the dscore computation formulas use the total activity count of all known users for a URL and the total activity count for all URLs. The activity counts of users who have missing values can therefore not be considered. In other words, the known users are treated differently for different attributes in that the attributes, for which the known users have provided information, are included in the URL\_dscore computation, whereas the  
20           attributes which are missing are not included in the dscore computation. However, inter-dependencies already existing between different attributes can provide more accurate predictions.

          Considering the known user as an unknown user for the missing attributes and predicting the missing attribute values in the same manner as for unknown users tends to be  
25           more accurate, since the attributes are computed from disclosed demographic information and the browsing activity of the known user. However, this process is computation-intensive and may be justified if the results have to be very accurate.

          The systems and methods described above may be implemented in digital electronic circuitry, or in computer hardware, firmware, software, or in combinations thereof.  
30           Apparatus of the invention can be implemented in a computer program product tangibly embodied in a machine-readable storage device for execution by a programmable



- 16 -

processor; and method steps of the invention can be performed by a programmable processor executing a program of instructions to perform functions of the invention by operating on input data and generating output. The invention can advantageously be implemented in one or more computer programs that are executable on a programmable  
5 system including at least one programmable processor coupled to receive data and instructions from, and to transmit data and instructions to, a data storage system, at least one input device, and at least one output device. Each computer program can be implemented in a high-level procedural or object-oriented programming language, or in assembly or machine language if desired; and in any case, the language can be a compiled  
10 or interpreted language. Suitable processors include, by way of example, both general and special purpose microprocessors. Generally, a processor will receive instructions and data from a read-only memory and/or a random access memory. Storage devices suitable for tangibly embodying computer program instructions and data include all forms of non-volatile memory, including by way of example semiconductor memory devices, such  
15 as EPROM, EEPROM, and flash memory devices; magnetic disks such as internal hard disks and removable disks; magneto-optical disks; and CD-ROM disks. Any of the foregoing can be supplemented by, or incorporated in, ASICs (application-specific integrated circuits).

To provide for interaction with a user, the invention can be implemented on a  
20 computer system having a display device such as a monitor or LCD screen for displaying information to the user and a keyboard and a pointing device such as a mouse or a trackball by which the user can provide input to the computer system. The computer system can be programmed to provide a graphical user interface through which computer programs interact with users.

25 While the invention has been disclosed in connection with certain illustrated embodiments shown and described in detail, various modifications and improvements thereon will become readily apparent to those skilled in the art. Moreover, the systems and methods described herein may be employed for a plurality of different applications including for generating profiles of unknown users on a computer network. Additionally,  
30 the systems and methods described herein may be employed for determining the success of a web site for attracting users of a selected profile, or demographic. Additionally, it will be

- 17 -

understood that the systems and methods described herein may be operated in a way to generated meaningful user profile information, without having to provide user identify information, such as user name, or address, as part of the profile. Accordingly, the spirit and scope of the invention is to be limited only by the following claims.

5

## Appendix

### 1. Notations

The following notations are used throughout the document:

$A_1, A_2, \dots, A_s$  represent the demographic attributes.

Each attribute  $A_i$  has a number of attribute values  $\{v_{i1}, v_{i2}, \dots, v_{iji}\}$  where  $j_i$  differs for each  $i$ .

Examples are:

$A_1$  = gender.  $v_{11}$  = male;  $v_{12}$  = female. In this case,  $j_i = 2$ .

$A_2$  = age.  $v_{21} = 2-11$ ;  $v_{22} = 12-17$ ;  $v_{23} = 18-24$ ;  $v_{24} = 25-34$ ;  $v_{25} = 35-44$ ;  $v_{26} = 45-54$ ;  $v_{27} = 55+$ . In this case,  $j_i = 7$ .

$K_1, K_2, \dots, K_t$  are the known users.

$U_1, U_2, \dots, U_z, \dots$  are the unknown users.

$R_1, R_2, \dots, R_y, \dots$  are the discriminating URLs.

$D_{yij}$  is the dscore of URL  $R_y$  for attribute value  $v_{ij}$  of attribute  $A_i$ .

A URL is called discriminating if its dscore for any attribute value differs from a typical distribution of that attribute value by a certain tunable threshold value, as described below in Section 2 of the Appendix.

### 2 Design of Dscore Equations for URLs

#### 2.1 Browsing activity of a known user

The computation of the Dscore for a URL will consider predominantly the browsing activity of known users rather than the number of known users. The browsing activity includes the duration for which the URL was browsed and the number of hits to the URL. The duration values should be significant enough to indicate the interest of the user in that URL. However, a scaling function is applied to the duration values to remove undue importance to a single visit having a large duration.

$$\text{Duration Count (DC)} = \lambda(d_{yt}) = \frac{\beta_1 d_{yt}}{\beta_1 d_{yt} + 1} \quad (\text{Equation 1})$$

where  $d_{yt}$  is the browsing duration.

$\beta_1$  is the tunable parameter for capping the duration for which a URL is visited by a user.

- 19 -

For every visit by a user, the duration count may be updated by adding the new duration  $\lambda(d_{y,i})$  count to the old duration count  $DC_{old}$  to give the new updated duration count  $DC_{new}$ :

$$DC_{new} = DC_{old} + \lambda(d_{y,i}) \quad (\text{Equation 2})$$

Each of multiple visits by a certain user may be treated as equivalent to a visit by another user having the same profile. However, a large number of multiple visits from the same user does not provide much additional information. The duration count is therefore scaled by a non-linear function which increases rapidly for a relatively small number of visits, but remains approximately constant for a relatively large number of visits. The result is called a final activity count  $AC$ .

The function is given by:

$$AC = \frac{\beta_2 * DC}{(\beta_2 * DC) + 1} \quad (\text{Equation 3})$$

$\beta_2$  is a tunable scaling parameter for capping the number of visits to a URL by a user and  $DC$  is the duration count determined above. The value of this function is less than one.

The effect of  $\beta_1$  and  $\beta_2$  on the activity count for different values of the number of visits and the duration of a visit can be seen from Tables A1-A4:

$\beta_1=0.1 \quad \beta_2=0.2$

Browsing duration					
		1	10	100	1000
Number of visits	1	0.018	0.091	0.154	0.165
	10	0.154	0.5	0.645	0.664
	100	0.645	0.909	0.948	0.952
	1000	0.948	0.994	0.005	0.995

Table A1

$\beta_1=0.1 \quad \beta_2=0.5$

Browsing duration					
		1	10	100	1000
Number of visits	1	0.043	0.2	0.312	0.331
	10	0.312	0.714	0.812	0.832
	100	0.820	0.962	0.978	0.980
	1000	0.974	0.996	0.998	0.998

Table A2

$\beta_1=0.2 \quad \beta_2=0.5$

- 20 -

	Browsing duration			
	1	10	100	1000
Number of visits	1	0.062	0.143	0.164
	10	0.4	0.625	0.662
	100	0.870	0.943	0.951
	1000	0.985	0.994	0.995

Table A3

 $\beta_1=0.5 \quad \beta_2=0.5$ 

	Browsing duration			
	1	10	100	1000
Number of visits	1	0.143	0.294	0.329
	10	0.625	0.806	0.831
	100	0.943	0.976	0.980
	1000	0.994	0.997	0.998

Table A4

As seen from the tables, a short duration and/or a small number of visits have reduced effect on activity count. The activity count increases with longer duration and/or the number of visits. The activity count increases less rapidly for long durations and a large number of visits. The effect of duration and number of visits on activity count can be adjusted by changing the values of  $\beta_1$  and  $\beta_2$ .

The change of a user's browsing pattern should be reflected in the dscores. For this, the activity counts of the users are "aged" to decrease the effect of old activity counts on the new dscore values.

To age the activity count  $AC$  of a known user, the activity count  $AC$  is first de-normalized. The de-normalized activity count  $AC'$  is calculated from the equation

$$AC' = \frac{AC}{\beta_2 - AC * \beta_2} \quad (\text{Equation 4})$$

wherein  $AC < 1$ .

For  $AC=1$ ,  $AC'$  is set to a very high value  $H$ .

The de-normalized activity count  $AC'$  is then aged to give  $AC'_{\text{aged}}$  by applying the aging formula

$$AC'_{\text{aged}} = AC' * 2^{-\frac{t-t_c}{2}} \quad (\text{Equation 5})$$

It will be appreciated that any exponentially decaying function can be used in lieu of equation 5.

The de-normalized aged activity count is then again normalized to give an aged normalized activity count

$$AC_{aged} = \frac{\beta_2 * AC_{aged}}{\beta_2 * AC_{aged} + 1} \quad (\text{Equation 6})$$

From the values of the activity count of known users computed above, the dscore  $D_{yij}$  for an attribute value  $v_{ij}$  for attribute  $A_i$  for a URL  $R_y$  can now be computed. In an exemplary embodiment, the following equation can be used to define the dscore  $D_{yij}$ :

$$D_{yij} = \frac{n_{yij} + t * \alpha * p_{ij}}{n_y + t * \alpha} \quad (\text{Equation 7})$$

wherein  $n_{yij}$  is the sum of activity counts for all known users  $K_l$  ( $l=1, \dots, t$ ) for whom  $A_i$  takes the value  $v_{ij}$  for URL  $R_y$ , i.e.

$$n_{yij} = \sum_{l=1}^t AC_{yl} * c \quad (\text{Equation 8})$$

For each attribute  $A_i$  and a known user  $K_p$ , only one of the  $v_{ij}$  will have the value 1, whereas the remaining  $v_{ij}$  for that attribute  $A_i$  and user  $K_p$  have the value zero. The parameter  $C$  is defined as

$$c = \begin{cases} 1, & \text{if } v_{ij} = 1 \text{ for user } K_l \\ 0 & \text{otherwise} \end{cases} \quad (\text{Equation 9})$$

The activity counts are calculated from equations 4 and 6, respectively.

$n_y$  is the sum of activity counts for all known users  $K_l$  for URL  $R_y$ , i.e.

$$n_y = \sum_{l=1}^t AC_{yl} \quad (\text{Equation 10})$$

In equation 7,  $t$  is the total number of known users;  $p_{ij}$  is the probability of known users having an attribute  $A_i$  with attribute values  $v_{ij}$ ;  $\alpha$  is an adjustable parameter to distribute the dscore proportionately between 0 and 1. It was experimentally observed that  $\alpha$  can suitably be set to  $1/n_y$ . However, other values of  $\alpha$  may further improve the dscore distribution. Also,  $n_{yij}$  will always be less than or equal to  $t * p_{ij}$ .

A URL is discriminating if its dscore for any attribute value  $v_{ij}$  differs from the normal distribution of a known user having that attribute value ( $p_{ij}$ ) by a certain defined threshold value  $\gamma$ , so that  $|D_{yij} - p_{ij}| > \gamma$ .

### 3. Design of Dscore Equations for Unknown Users

According to our problem specification, we want to predict a demographic profile for an unknown user. The demographic profile is represented in the form of a dscore for each attribute value for the unknown user. This dscore value depends on following factors:

- Dscores of the discriminating URLs being browsed in the current session
- Old dscore of the unknown user
- Browsing activity of the unknown user for the current session

The dscore of an unknown user will be such that if the browsing activity is sufficiently large the dscores are a combination of the URLs browsed. But if the browsing activity is insignificant the dscore will weigh down to that of a typical user.

### 3.1 Browsing Activity for an Unknown User

The duration counts for unknown user visits are computed by a method identical to that used for computing duration counts for known users using equation 2.

Like the de-normalized activity count  $AC'$  for known users, the duration count for unknown user is aged to using a formula similar to equation 5:

$$DC_{aged} = DC * 2^{-\left(\frac{t-t_c}{2}\right)} \quad (\text{Equation 11})$$

The dscores for each demographic attribute  $A_i$  of the unknown user can now be computed from the duration count of the unknown user and the dscores of the URL obtained from equation 7. To predict the dscores for each demographic attribute value of the unknown user, the discriminating ability of the browsed URL and the duration count obtained from the browsing activity of the unknown user should be suitably weighted, for example, by the following function  $f(x)$ :

$$f(x) = \frac{(e^{c(x-p)} - e^{-c(x-p)})}{2} \quad (\text{Equation 12})$$

wherein  $c$  is an adjustable parameter;  $x$  is a variable, in this case the dscore of the URL browsed; and  $p$  is the probability of a known user having an attribute value  $v_{ij}$  in the original distribution.

The function  $f(x)$  assumes a high value for those values of  $x$  that are far away from the original probability distribution  $p$ , and a low value for those values of  $x$  that are close to  $p$ . Hence the dscores of the URLs that are more discriminating are scaled by a larger amount. As a result, the dscore values of the unknown user reflects the discriminating measure  $(x-p)$  of a URL.

To compute the dscores for all URLs that an unknown user browses in a session, the value of  $f(dscore)$  is first computed for each attribute value. The so computed values of  $f(dscore)$  are then weighted and averaged, with the weights being the duration counts

for the URLs. Finally, the inverse function  $f'$  of the resulting value is calculated to provide the new dscore value for the unknown user:

$$D_{new} = f^{-1}\left(\frac{f(D_1) * DC_1 + f(D_2) * DC_2 + \dots}{DC_1 + DC_2 + \dots}\right) \quad (\text{Equation 13})$$

where  $D_i$  is the dscore of a URL  $R_i$  browsed by the unknown user in that session,  
 $DC_i$  is the duration count for URL  $R_i$ ,  
 $D_2$  is the dscore of a URL  $R_2$  browsed by the unknown user in that session,  
 $DC_2$  is the duration count for URL  $R_2$ ,  
 and so on.

For example, an unknown user browses URLs  $R_1$  and  $R_2$  with  $DC_1=DC_2=1$ . Also, the dscores for  $R_1$  and  $R_2$  for a particular attribute value are assumed to be 0.9 and 0.6, respectively, and the probability  $p$  of a known user having an attribute value  $v_{ij}$  in the original distribution is assumed to be 0.

The following dscores are calculated from the above values:

$c$	$f(R_1)$	$f(R_2)$	Avg. of $f(R_1)$ and $f(R_2) = DA$	$dscore_{calc} = f'(DA)$
1	1.03	0.64	0.83	0.755
2	2.94	1.51	2.23	0.775
5	45.00	10.02	27.51	0.795
10	4051	201.7	2126	0.835
20	$32.8 * 10^6$	81377	$16.4 * 10^6$	0.865

Table A5

As seen from the values of the original dscore of  $R_1$  and  $R_2$ ,  $R_1$  is more discriminating than  $R_2$ . Consequently, the  $dscore_{calc}$  of the unknown user is more heavily weighted towards  $R_1$ . Moreover, as the value of  $c$  increases, the  $dscore_{calc}$  of an unknown user is pulled more towards the dscore of the URL that is more discriminating.

The newly calculated  $dscore_{calc}$  for a session of the unknown user may be merged with existing dscores for the unknown user. Furthermore, the duration counts also need to be updated. The process therefor is as follows:

For an attribute value  $v_{ij}$ , let the old dscore of the unknown user  $U$  be  $D_{old}$  and the old duration count  $DC_{old}$ , both calculated at a time  $t_c$ . In the next session, at time  $t$ , a new dscore  $D_{new}$  and a new duration count  $DC_{new}$  is calculated for the same unknown user  $U$ , using equation 13. The old duration count  $DC_{old}$  and the old dscore  $D_{old}$  are first aged according to equation 11 to give an aged duration count  $DC_{old\_Aged}$  and an aged dscore  $D_{old\_Aged}$ . The new dscore is then calculated using equation 13 as follows:

$$D_{new} = f^{-1}\left(\frac{f(D_{old\_Aged}) * DC_{old\_Aged} + f(D_1) * DC_1}{DC_{old\_Aged} + DC_1 + k}\right) \quad (\text{Equation 15})$$



- 24 -

Initially,  $D_{old}$  and  $DC_{old}$  are zero.  $k$  is a parameter related to a significance of a user activity. Since  $f'(0) = p$ , where  $p$  is the probability of a known user having an attribute value  $v_{ij}$  in the original distribution, the dscore predicted for the unknown user will substantially be equal to a typical profile, if the activity count of the unknown user is small compared to  $k$ . Otherwise, the dscores are those predicted by the URLs browsed. The new duration count  $DC_{new}$  is obtained by adding  $DC_{old\_Aged}$  to  $D_1$ .

Like the dscores of known users, the dscores of unknown users may also be aged. According to equation 15, a current unknown user dscore  $D_{curr}$  and a current duration count  $DC_{curr}$  is related to the sum  $X$  of the products of old  $f$ (dscores) and duration counts  $DC$  by the following relation computed at time  $t_c$ :

$$D_{curr} = f^{-1}\left(\frac{X}{DC_{curr} + k}\right) \quad (\text{Equation 16})$$

which is equivalent to

$$X = f(D_{curr}) * (DC_{curr} + k). \quad (\text{Equation 16a})$$

At a future time  $t$ ,  $D_{curr}$  ages to  $D_{curr\_Aged}$ . Assuming that both  $DC_{curr}$  and  $X$  age according to equation 5 with an aging time  $\lambda$ , then the aged dscore  $D_{curr\_Aged}$  for unknown users can be calculated as:

$$D_{curr\_Aged} = f^{-1}\left(\frac{X * 2^{\frac{t-t_c}{\lambda}}}{DC_{curr} * 2^{\frac{t-t_c}{\lambda}} + k}\right) \quad (\text{Equation 17})$$

or by substituting equation 16a

$$D_{curr\_Aged} = f^{-1}\left(\frac{f(D_{curr}) * (DC_{curr} + k) * 2^{\frac{t-t_c}{\lambda}}}{DC_{curr} * 2^{\frac{t-t_c}{\lambda}} + k}\right) \quad (\text{Equation 18})$$

#### 4. Configuration Options

The following configuration parameters were found to provide satisfactory results for calculating dscores:

Parameter	Set by	Values	Default value	Description
$c$	User	1, 1.5, 2, 2.5, ..., 20	10	Used in URL dscore calculations for weighting the discriminating behavior of a URL
$\beta_1$	User	0, ..., 1	0.1	Tunable parameter for capping the duration for which a URL is visited by a user
$\beta_2$	User	0, ..., 1	0.5	Tunable parameter for capping the number of visits to a URL by a user
$\gamma$	User	0, ..., 1	0.1	Parameter to decide if a URL is discriminating

- 25 -

1. In a computer network formed of a communication channel and a plurality of digital processors coupled to the communication channel for communication thereon and providing information to a user, a method for generating a demographic profile for an unknown user comprising:

5                    recording computer activity of the unknown user in response to the information provided to the user by at least one of the digital processors, and  
                     combining the recorded computer activity of the unknown user with a computer demographic score of the at least one of the digital processors, the computer demographic score being based on demographic information obtained from known users, to  
10                    generate the demographic profile of the unknown user.

2. A process for generating a demographic profile of an unknown user accessing a server, the server having a server profile, the process comprising:

                     recording computer activity by the unknown user in response to information  
15                    provided by the server;

                     if the recorded computer activity by the unknown user is greater than a predetermined activity value, combining the recorded computer activity by the unknown user with the server profile to form the unknown user demographic profile, and

                     if the recorded computer activity by the unknown user is less than a  
20                    predetermined activity value, setting the unknown user demographic profile equal to the server profile.

3. The process of claim 2, wherein a weighting function is applied to the recorded computer activity by the unknown user based on a duration of the computer activity.

25

4. The process of claim 3, wherein the applied weighting function is selected to reduce the significance of a computer activity having a long duration.

5. The process of claim 2, wherein the unknown user demographic profile formed  
30                    during a first session with the server is retained in a history table and merged with the unknown user demographic profile formed during a subsequent second session.

- 26 -

6. The process of claim 5, wherein the unknown user demographic profile formed in the first session with the server is aged before being merged with the unknown user demographic profile formed during the subsequent second session.

5

7. The process of claim 2, wherein at least a portion of the server profiles are stored in memory cache.

8. The process of claim 7, wherein the memory cache is a hash table.

10

9. The process of claim 2, wherein the unknown user accesses at least two servers and the unknown user demographic profile is formed by merging the user demographic profiles from all accessed servers that have a server profile.

15 10. The process of claim 2, wherein the server profile is updated in response to a computer activity by a known user.

11. A process for generating a user demographic profile of an unknown user accessing at least one server, comprising:

20 identifying a user accessing the at least one server and recording user activity on the server;

based on the user identification, determining if the user is an unknown user or a known user;

for an unknown user:

25 monitoring at least a duration of the user activity and assigning a demographic score to the unknown user based on the monitored user activity and a server profile of the at least one server accessed by the unknown user;

combining the demographic score with an existing demographic score of the unknown user; and

30 setting the demographic profile of the unknown user equal to the combined demographic score.

- 27 -

12. The process of claim 12, combining comprises merging the demographic scores obtained from the same server and from another server accessed by the unknown user.
- 5 13. The process of claim 12, further comprising for a known user:  
monitoring at least a duration of the known user activity and updating a browsing record of the known user;  
updating the server profile in response to the updated browsing record; and  
using the updated server profile to compute the demographic score of the  
10 unknown user.
14. The process of claim 12, wherein the user is identified based on a cookie returned by the at least one server.
- 15 15. In a computer network formed of a communication channel and a plurality of computers coupled to the communication channel for communication thereon, a computer program, residing on a computer-readable medium, comprising instructions for causing a computer to:  
record a computer activity of a user responding to information provided by  
20 at least one of the computers;  
identifying the user as one of a known and an unknown user for the at least one computer;  
for an unknown user, compare the computer activity with a predetermined activity and assign to the unknown user a demographic score which is based on the  
25 computer activity and a computer demographic profile characteristic of the computer, if the computer activity exceeds the predetermined activity, and on the computer demographic profile alone, if the computer activity is less than the predetermined activity;  
combine the demographic score with another existing demographic score for the same unknown user generated during a previous session by the unknown user with the same  
30 computer or with another computer; and

- 28 -

provide from the combination of the demographic scores a user demographic profile of the unknown user.

16. The process of claim 15, wherein the computer demographic profile is  
5 generated from the computer activity of the known users.

17. The process of claim 15, wherein the computer activity is generated in response to information provided to the known or unknown user by the computer.

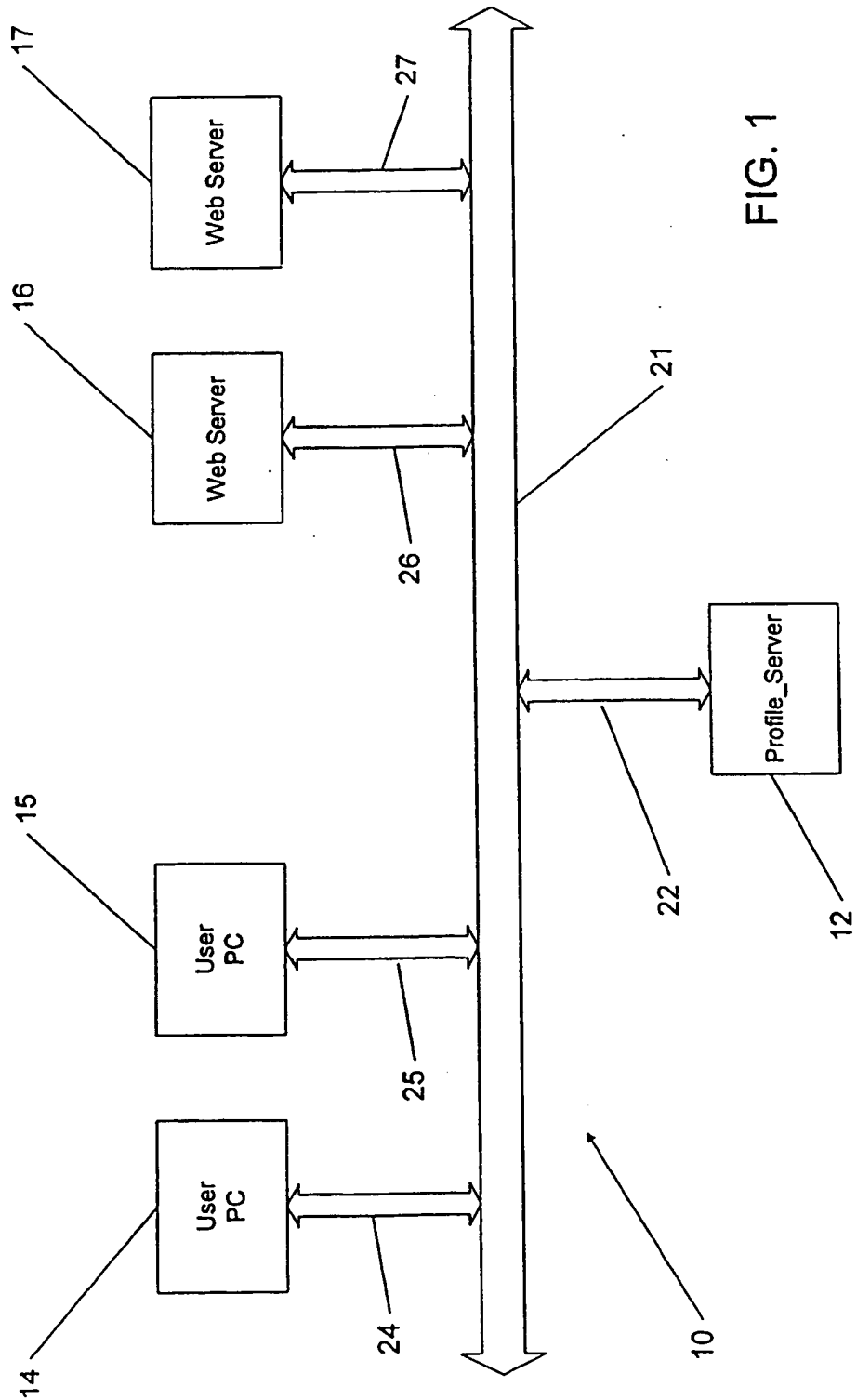


FIG. 1

2 / 13

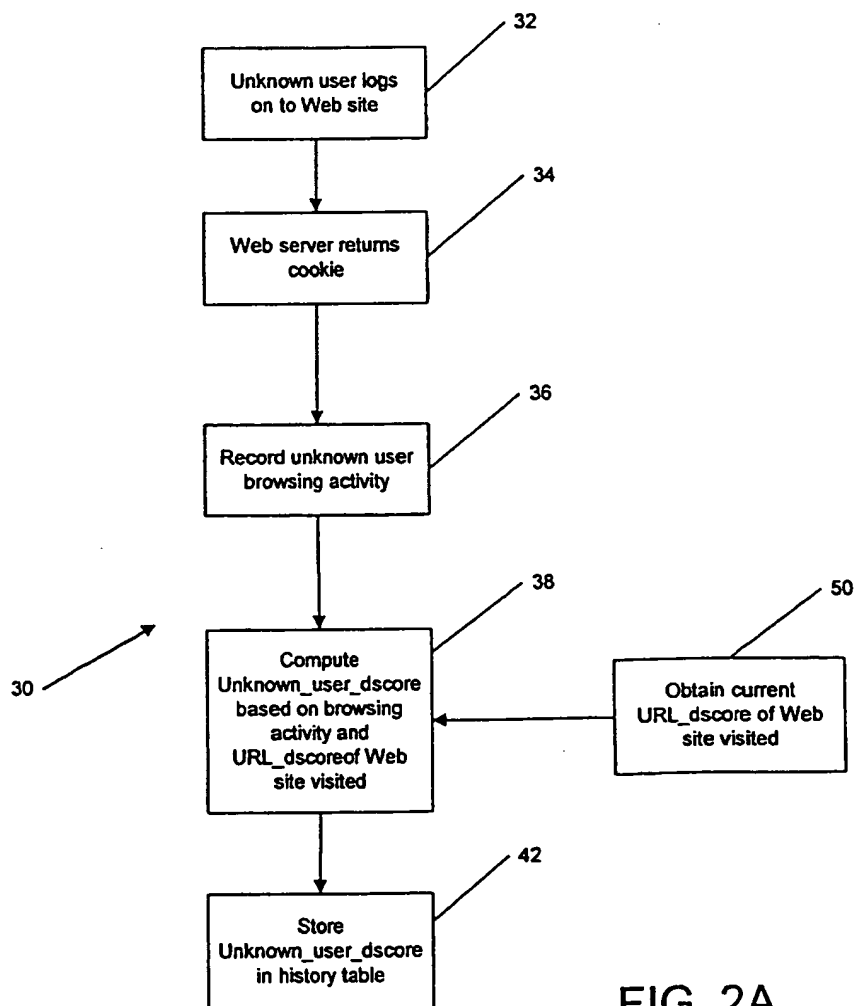


FIG. 2A

3 / 13

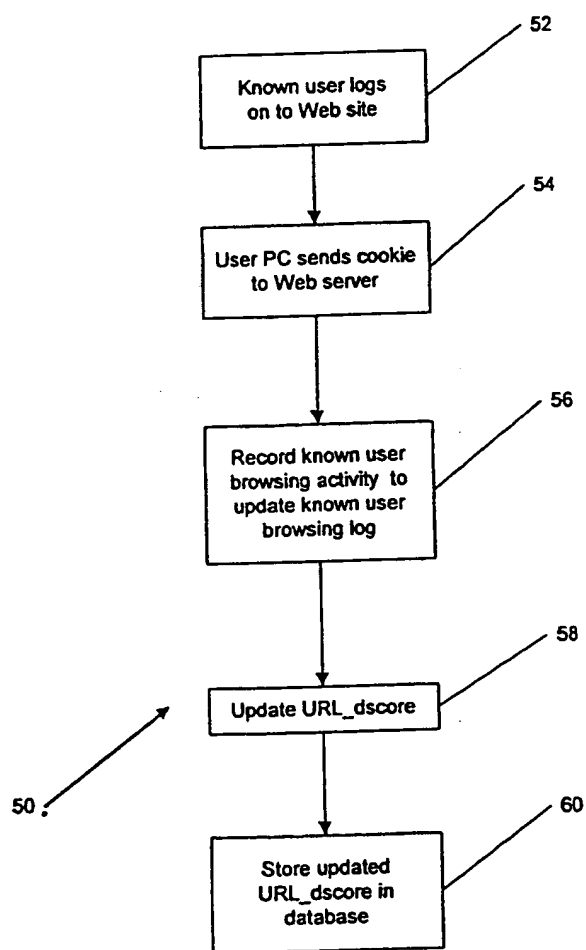


FIG. 2B



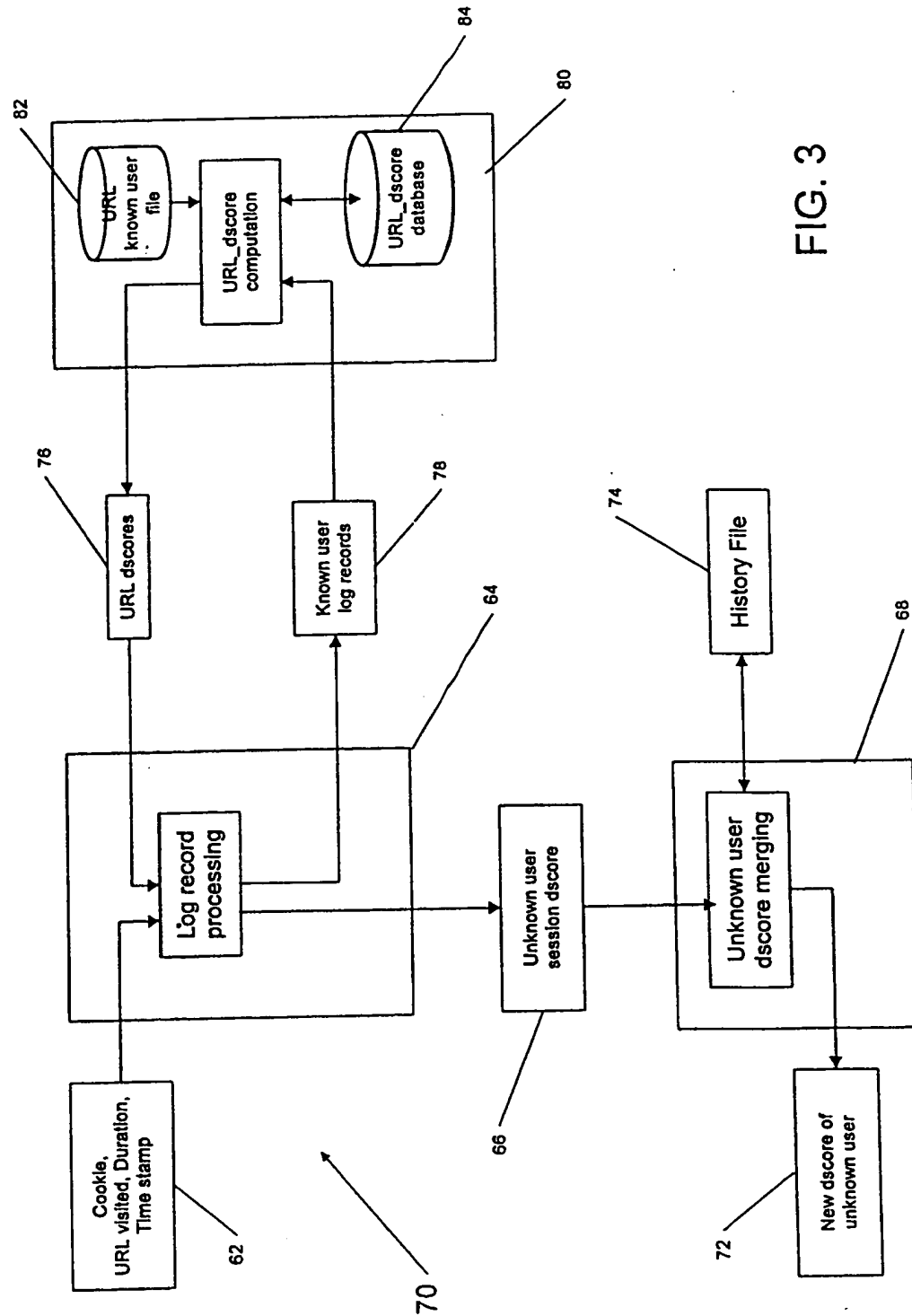


FIG. 3

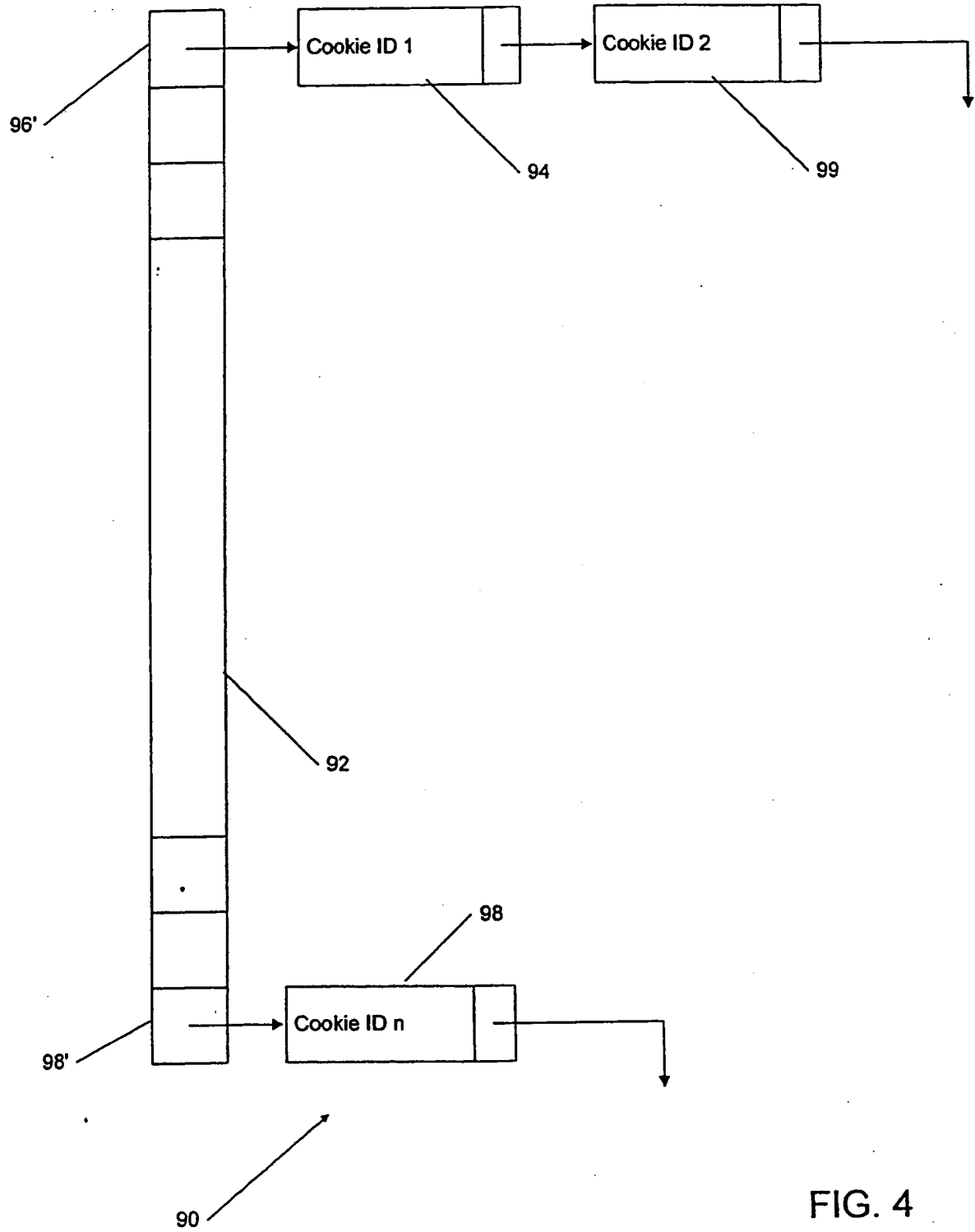


FIG. 4

6 / 13

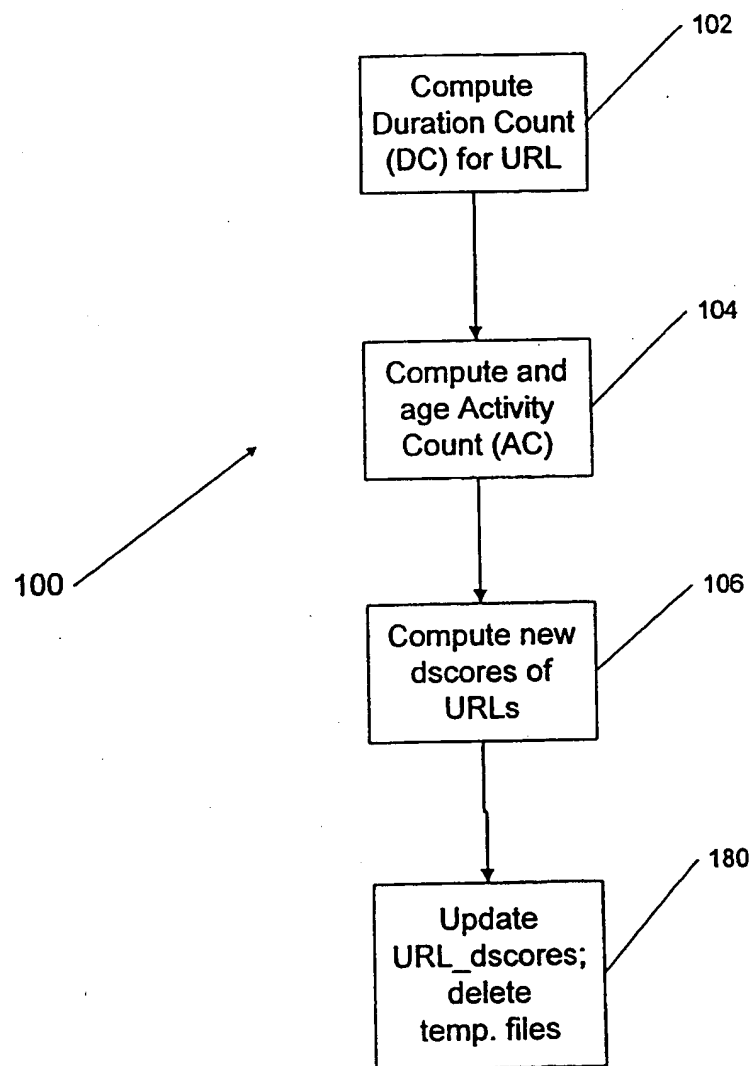
Known Users

FIG. 5

7 / 13

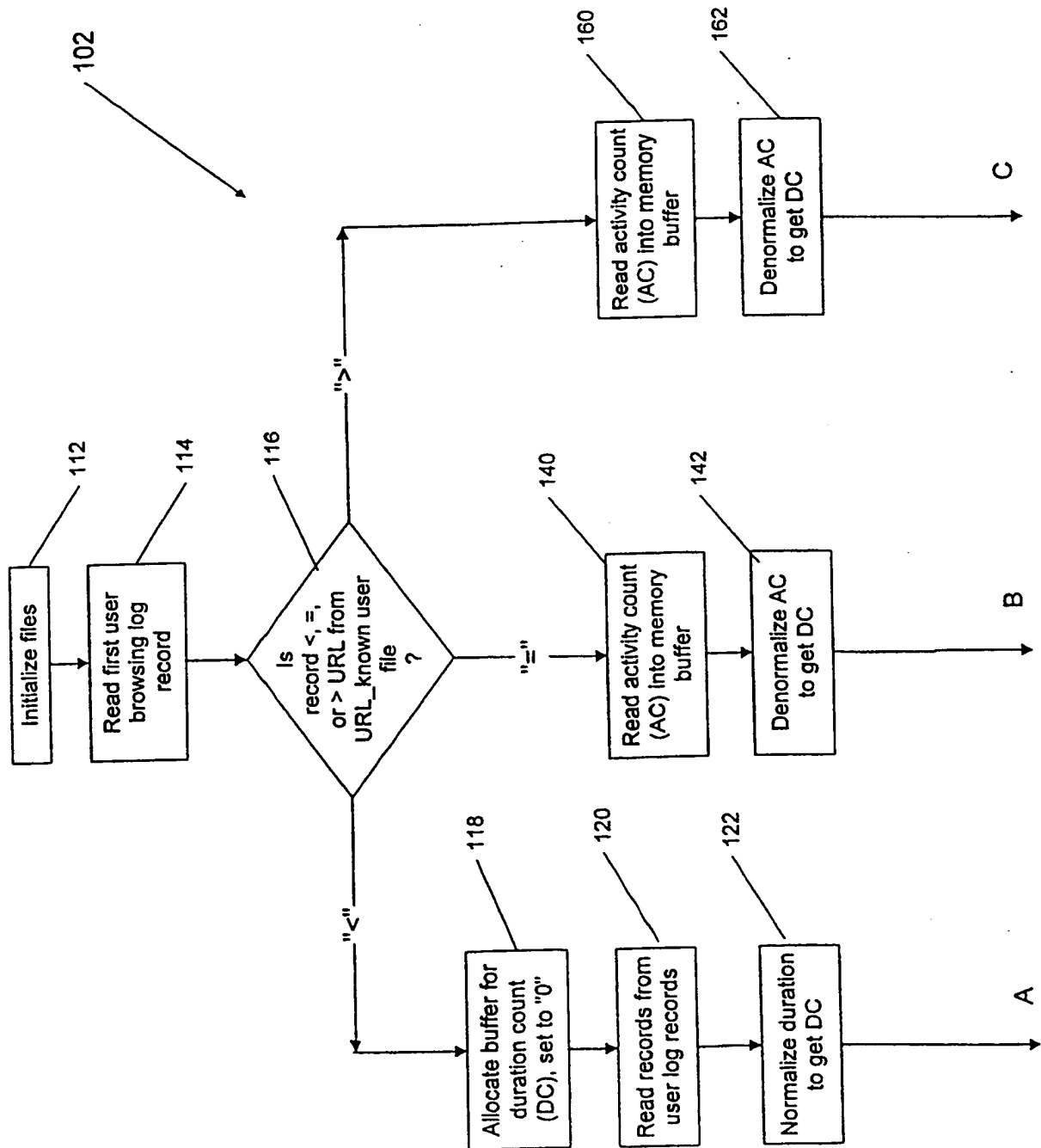


FIG. 6A

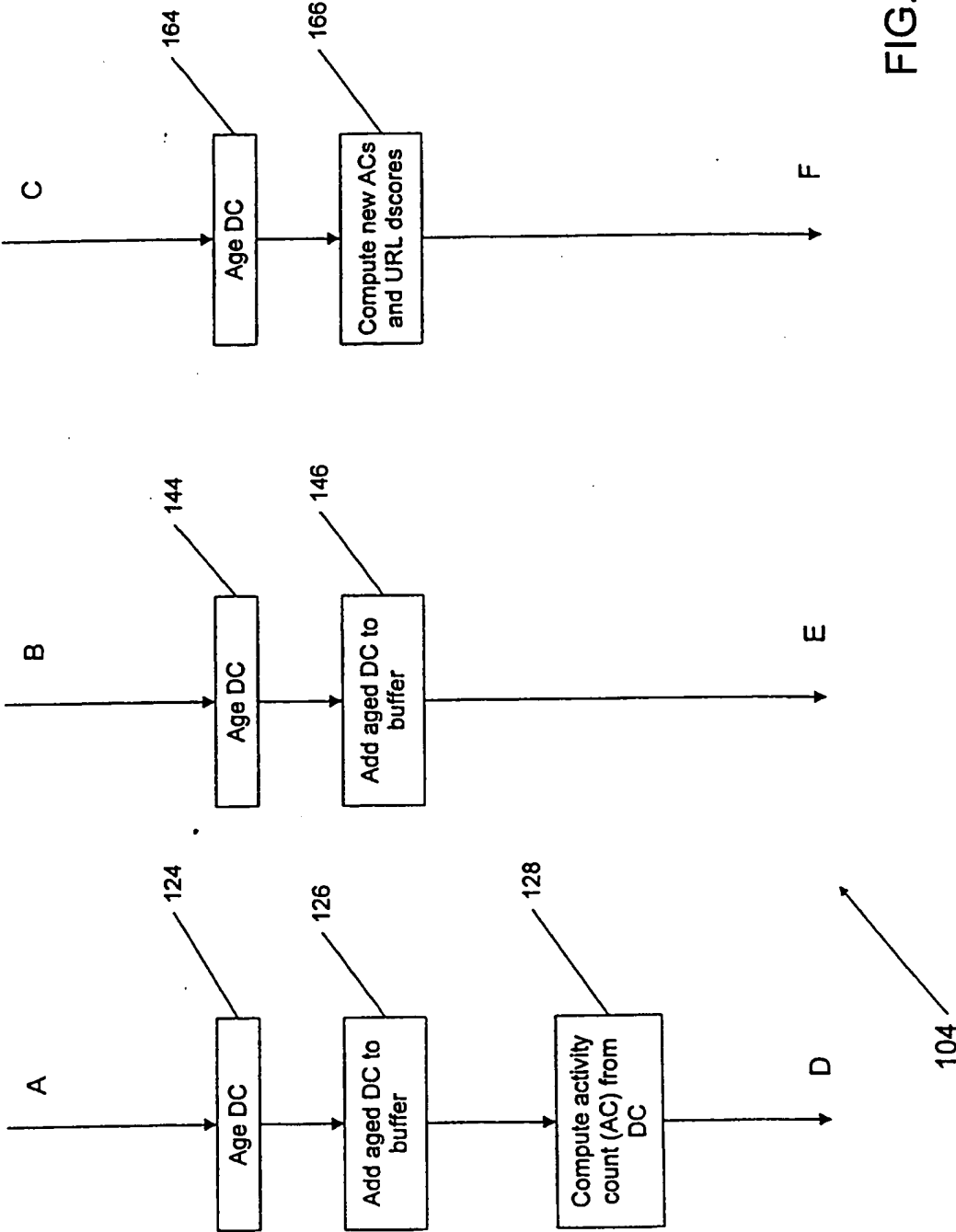
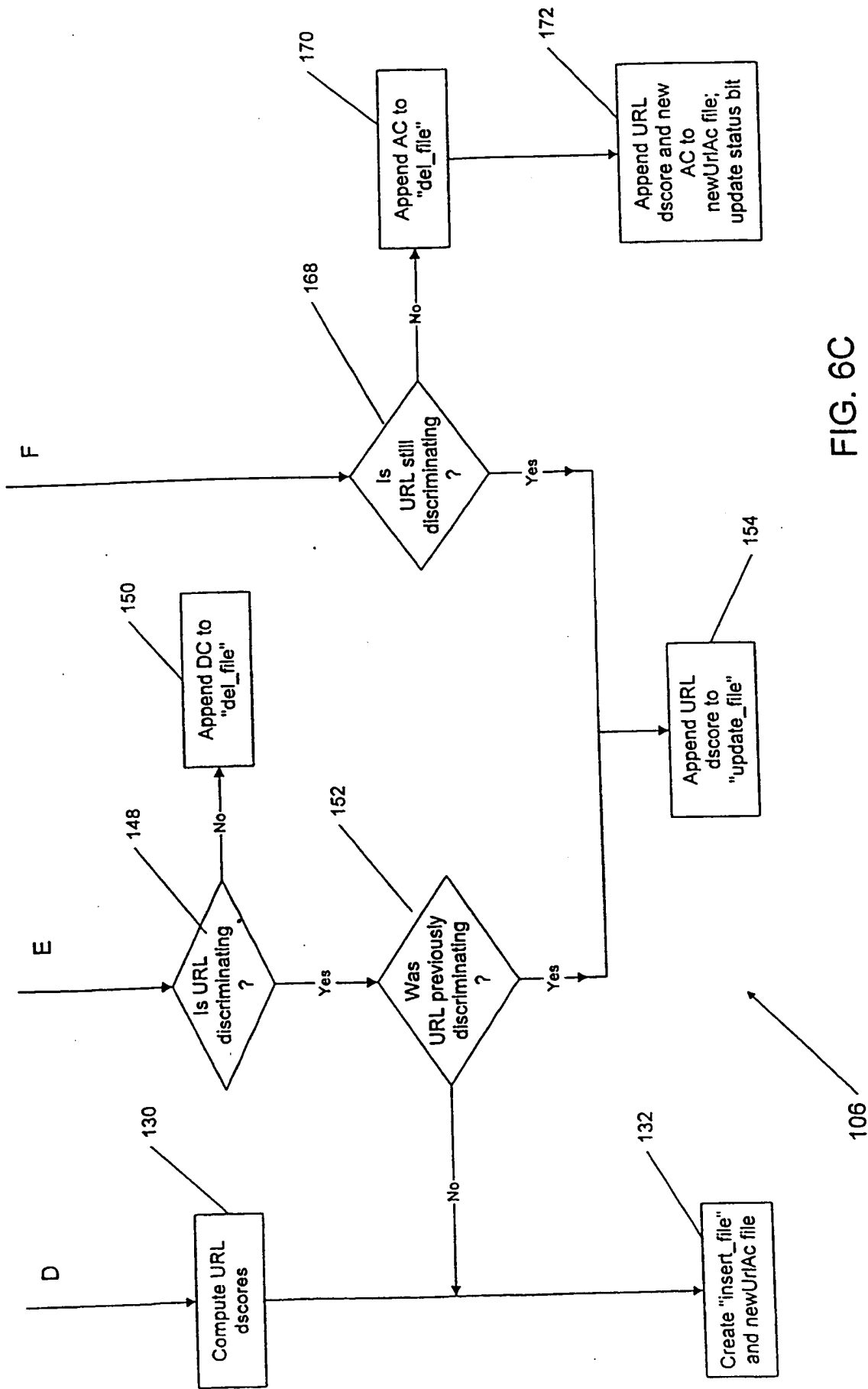


FIG. 6B



10 / 13

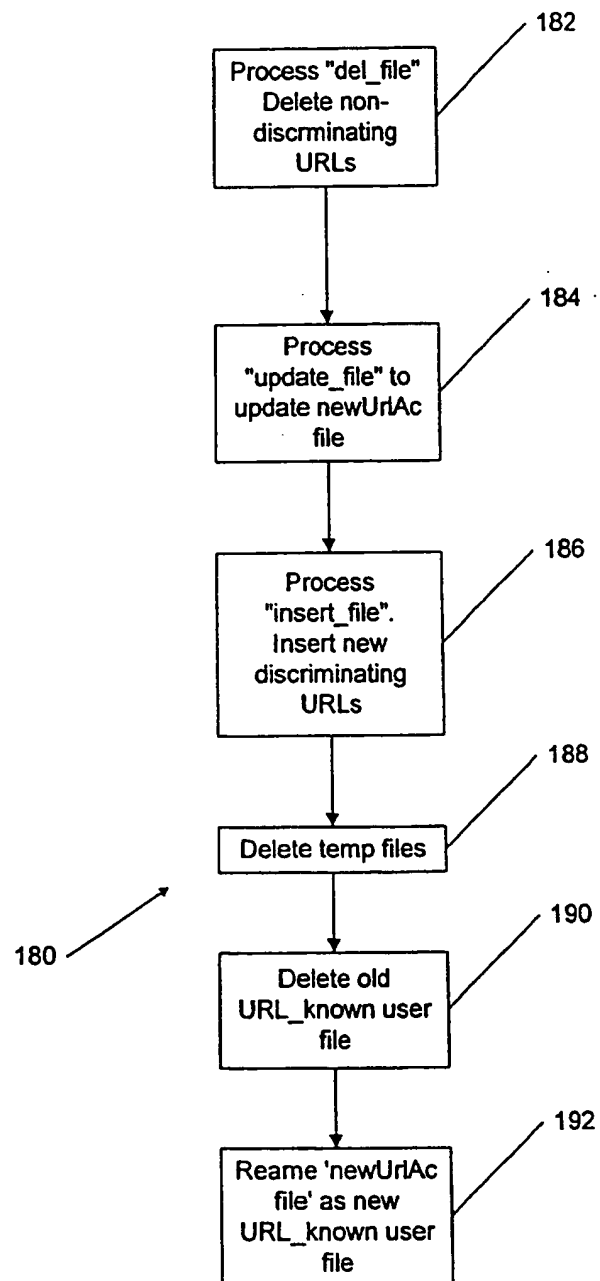


FIG. 7

11 / 13

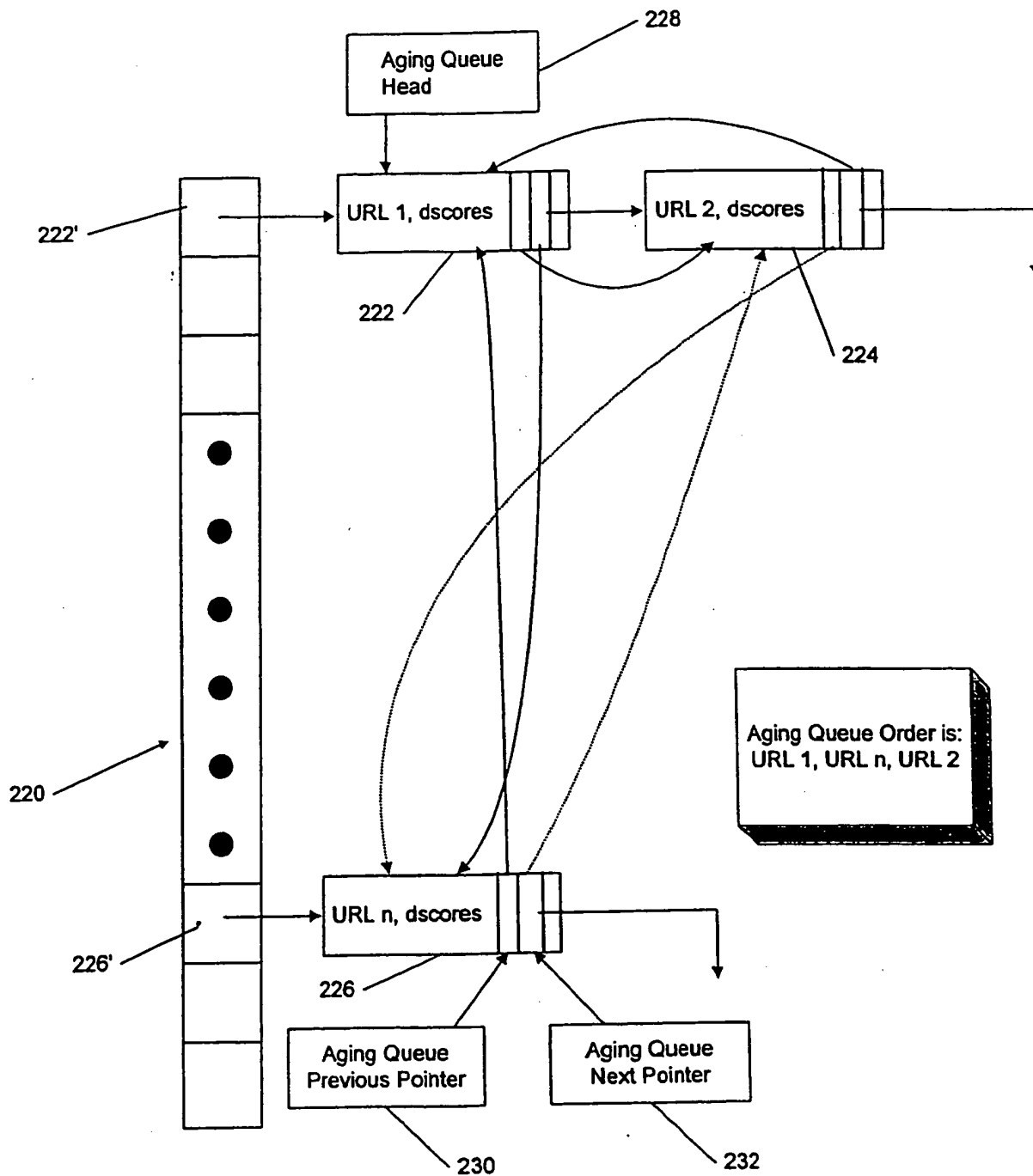


FIG. 8A



12 / 13

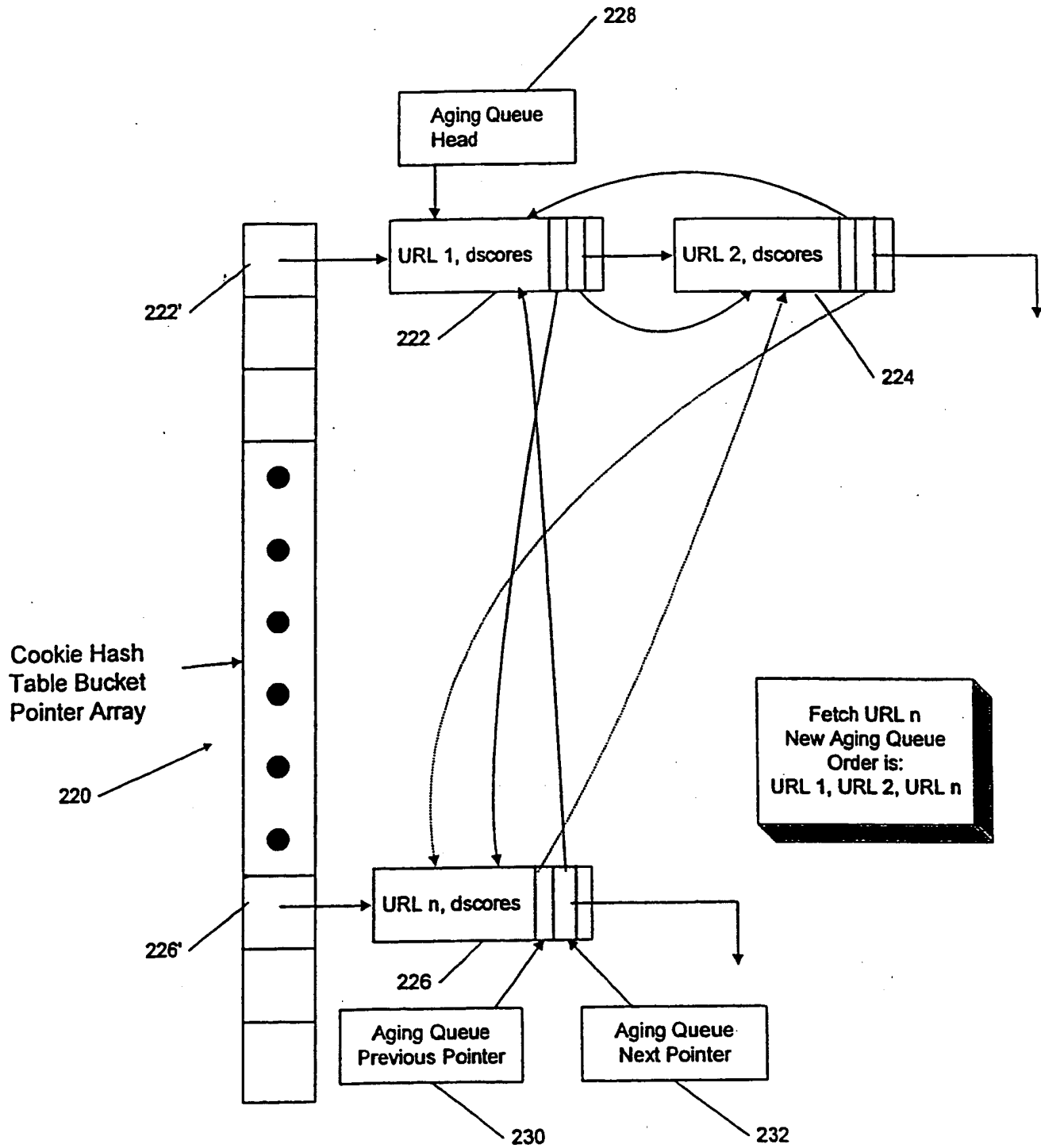


FIG. 8B

13 / 13

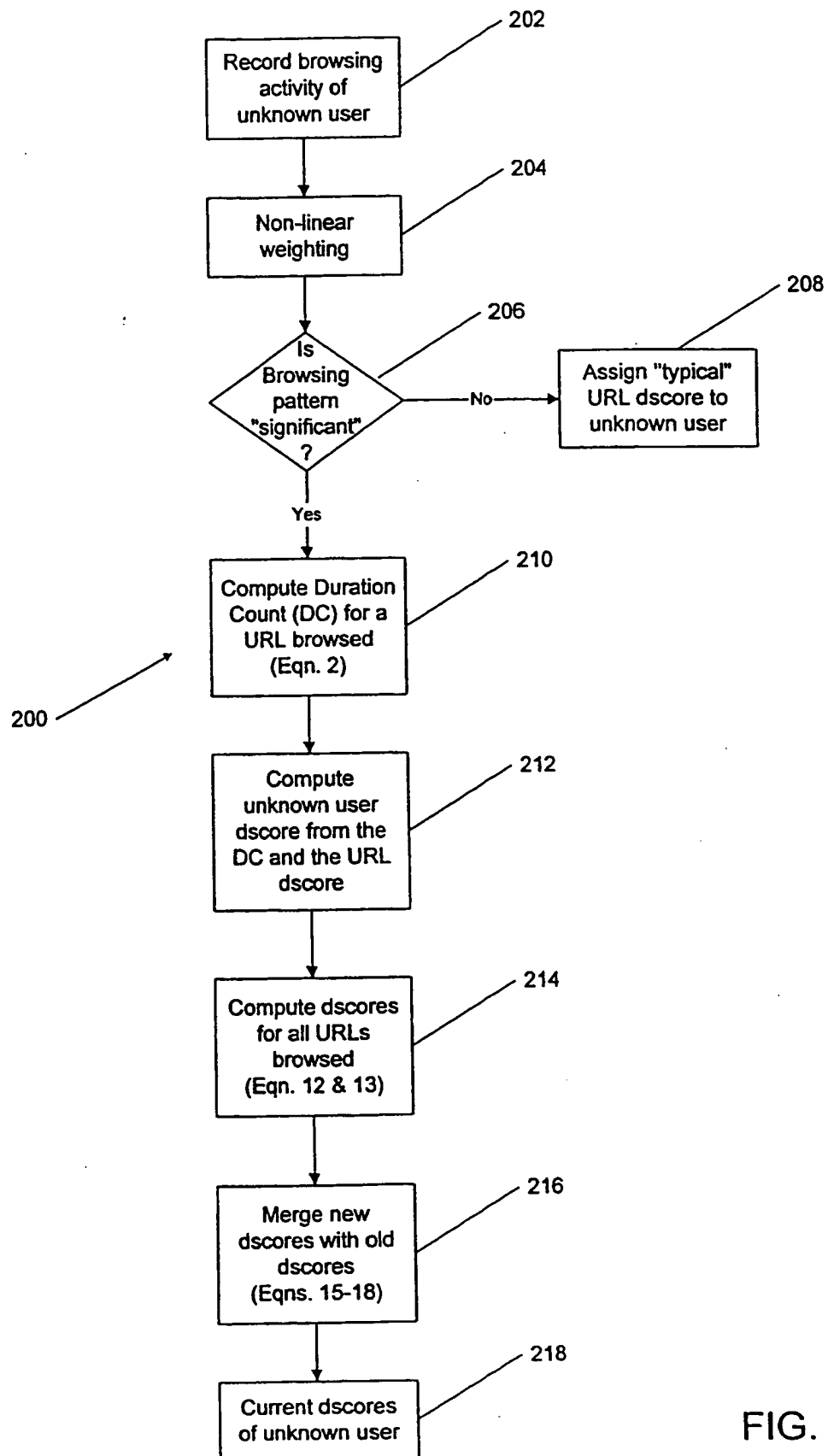
Unknown Users

FIG. 9

# PATENT COOPERATION TREATY

# PCT

## DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT

(PCT Article 17(2)(a), Rules 13ter.1(c) and Rule 39)


Applicant's or agent's file reference <b>ENK-003.25</b>	IMPORTANT DECLARATION	Date of mailing(day/month/year) <b>07/06/2001</b>
International application No. <b>PCT/US 01/ 03214</b>	International filing date(day/month/year) <b>31/01/2001</b>	(Earliest) Priority date(day/month/year) <b>31/01/2000</b>
International Patent Classification (IPC) or both national classification and IPC <b>G06F17/60B2</b>		
Applicant <b>ENGAGE TECHNOLOGIES</b>		

This International Searching Authority hereby declares, according to Article 17(2)(a), that **no international search report will be established** on the international application for the reasons indicated below

1. ☒ The subject matter of the international application relates to:
  - a. ☐ scientific theories.
  - b. ☐ mathematical theories
  - c. ☐ plant varieties.
  - d. ☐ animal varieties.
  - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
  - f. ☒ schemes, rules or methods of doing business.
  - g. ☐ schemes, rules or methods of performing purely mental acts.
  - h. ☐ schemes, rules or methods of playing games.
  - i. ☐ methods for treatment of the human body by surgery or therapy.
  - j. ☐ methods for treatment of the animal body by surgery or therapy.
  - k. ☐ diagnostic methods practised on the human or animal body.
  - l. ☐ mere presentations of information.
  - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.
  
2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:
 

☐ the description
 ☐ the claims
 ☐ the drawings
  
3. ☐ The failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions prevents a meaningful search from being carried out:
 

☐ the written form has not been furnished or does not comply with the standard.
 ☐ the computer readable form has not been furnished or does not comply with the standard.
  
4. Further comments:

Name and mailing address of the International Searching Authority  

 European Patent Office, P.B. 5818 Patentlaan 2  
 NL-2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

**Mar'a Rodr'guez N'voa**

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 203

The subject-matter claimed in claim 1 falls under the provisions of Article 17(2)(a)(i) and Rule 39.1(iii), PCT, such subject-matter relating to a method of doing business.

Claims 2-17 relate to a conventional process or computer program for performing the business method of claim 1. Although these claims do not literally belong to the method category, they essentially claim protection for the same commercial effect as the method claims. The International Searching Authority considers that searching this subject-matter would serve no useful purpose. It is not at present apparent how the subject-matter of the present claims may be considered defensible in any subsequent examination phase in front of the EPO as International Preliminary Examining Authority with regard to the provisions of Article 33(1) PCT (novelty, inventive step); see also Guidelines B-VII, 1-6).

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.5), should the problems which led to the Article 17(2) declaration be overcome.

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**